REMARKS

The Examiner has rejected claims 1-4, 7-10 and 13 under 35 U.S.C. 103(a) as being unpatentable U.S. Patent 5,918,223 to Blum et al. in view of the Sheirer et al. article "Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator", Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97), Vol 2, p1131-1134). The Examiner has further rejected claims 5 and 6 under 35 U.S.C. 103(a) as being unpatentable over Blum et al. in view of Sheirer et al., and further in view of the article "Quantitative Effects of Global Tempo on Expressive Timing in Music Performance: Some Perceptual Evidence" by B.H. Repp.

The Blum et al. patent discloses a method and article of manufacture for content-based analysis, storage, retrieval, and segmentation of audio information.

The Examiner has indicated that Blum et al. discloses the claim limitations "analyzing said audio signal to extract at least one predetermined audio feature" (Abstract lines 1-4, analysis... of audio files produces a set of feature vectors), "performing a frequency analysis on a set of values of said audio feature at different time instances resulting in a [power spectrum] ***magnitude spectrum*** of said extracted predetermined audio feature" (Col 15 lines 43-44, bass spectrum, which represents the bass trajectory at different time instances, is subjected to an FFT), "deriving at least one further audio feature representing a temporal behavior of said extracted predetermined audio feature by parameterizing said

[power spectrum] ***magnitude spectrum*** (Col 15 lines 50-60, beats detected from magnitude peaks representing a temporal behavior) and "classifying said audio signal based on said further audio feature" (Col 21 lines 53-65, the signal is classified into categories using statistical measures derived from the feature vectors), with the exception that, as noted, Blum et al. discloses "magnitude spectrum" as opposed to "power spectrum".

The Scheirer et al. article discloses a real-time computer system capable of distinguishing speech signals from music signals over a wide range of digital audio input. In the article, Scheirer et al. identifies 13 features which have been evaluated for use in the system, of which the following features are used in the system "4 Hz modulation energy", "Percentage of 'Low-Energy' Frames", "Spectral Rolloff Point", "Spectral Centroid", "Spectral 'Flux' (Delta Spectrum Magnitude)", "Zero-Crossing Rate", "Cepstrum Resynthesis Residual Magnitude", and "Pulse metric". Of these features, only the "Spectral Rolloff Point mentions "power spectrum", i.e., "The 95th percentile of the power spectral distribution. This measure distinguishes voiced from unvoiced speech-unvoiced speech has a high proportion of energy contained in the high-frequency range of the spectrum, where most of the energy for unvoiced speech and music is contained in lower bands. This is a measure of the "skewness" of the spectral shape-the value is higher for right-skewed distributions".

It is unknown to Applicants how this disclosure of Scheirer et al. relates to the subject invention, or how the

Examiner is taking the mere mention of "power spectral" from the entire disclosure of Scheirer et al. and deeming that this is equivalent to "magnitude spectrum" as disclosed in Blum et al., such that the terms may be interchanged.

Independent claims 1 and 8 include the limitation "performing a frequency analysis on a set of values of said extracted predetermined audio feature at different time instances resulting in a power spectrum of said extracted predetermined audio feature". Applicants submit that this is neither disclosed nor suggested by Blum et al. In particular, Blum et al. states, at col. 15, lines 42-49:

> "If the rhythm option is chosen, an FFT is performed on the bass trajectory. This yields a spectrum whose x-axis measures distances in time, and whose peaks indicate the most frequent separation in time between bass notes. For example, if the bass drum usually plays on the first beat of the measure, the time separation corresponding to one measure will show up as a peak."

It should be apparent from the above that the frequency analysis done by Blum et al. does not result in a power spectrum of the extracted predetermined audio feature.

Further, since Blum et al. does not disclose or suggest "a power spectrum of the extracted predetermined audio feature", then surely, Blum et al. neither discloses nor suggests "deriving at least one further audio feature representing a temporal behavior of said extracted predetermined audio feature by parameterizing said power spectrum".

In regard to claim 5, the Examiner states:

> "Blum discloses the deriving step comprises the steps of: calculating an average value of said set of values of said extracted predetermined audio feature at different time instances (Col 15 lines 43-44, taking an FFT produces frequency coefficients, the lowest of which is the DC value, or time average, of the signal for the given frame); defining at least one frequency band (Col 15 lines 43-44, taking an FFT defines at least one frequency bin); calculating the amount of energy within said frequency band from said frequency analysis (Col 15 lines 43-44, taking an FFT calculates coefficients representative of the amount of energy in each frequency bin); and defining said further audio feature as said amount of energy (Col 15 lines 44-46)."

The noted section of Blum et al. states:

> "If the rhythm options is chosen, an FFT is performed on the bass trajectory. This yields a spectrum whose x-axis measures distances in time, and whose peaks indicate the most frequent separation in time between bass notes."

It is unknown to Applicants how the Examiner was able to formulate the above analysis from just this disclosure in Blum et al., where none of the steps indicated in claim 5 appear to be disclosed.

The Examiner then concedes that "Blum and Shierer do not specifically mention defining said further audio feature as said amount of energy divided by said average value." The Examiner then states "Repp discloses defining a audio feature as an amount of energy divided by an average value (p41, calculation of relative modulation depth requires dividing energy by an average value)."

It is unknown to Applicants where in Repp the Examiner finds the concept that the term "relative modulation depth" (RMD) requires dividing energy by an average value. Rather, Repp, which states on page 41 "relative modulation depth (RMD), which will be

defined more precisely later." On page 43, Repp states "The measure of RMD is the slope of the regression line divided by the correlation."

In view of the above, Applicants believe that the subject invention, as claimed, is not rendered obvious by the prior art, either individually or collectively, and as such, is patentable thereover.

Applicants believes that this application, containing claims 1-10 and 13, is now in condition for allowance and such action is respectfully requested.

Respectfully submitted,

by /Edward W. Goodman/
Edward W. Goodman, Reg. 28,613
Attorney
Tel.: 914-333-9611